Red Participativa de Ciencias

S.- Gil y E. Rodríguez

www.cienciaredcreativa.org

Métodos cualitativos de análisis gráfico

Importancia de la representación gráfica

El análisis de datos, es una actividad que ha tenido un gran desarrollo en los últimos tiempos y es una actividad muy útil y común a un gran número de disciplinas académicas, actividades científicas, económicas y sociales. Es realmente útil que los datos a estudiar se presenten en un gráfico, pues aquí queda concentrada la información para su apreciación y análisis, además los gráficos muchas veces sugieren tendencias y relaciones entre las variables que se estudian, que muchas veces resulta muy difícil de detectarlas de otro modo. En la mayoría de los casos un gráfico es más útil que una tabla de valores, especialmente en los casos en que:[1]

- ✓ los estudios se llevan a cabo analizando una variable Y en función de otra X, y se quiere interpretar o determinar la relación funcional entre ellas. Por ejemplo: consumo de electricidad en función de la temperatura, ingreso medio de una persona en función de los años de educación formal, medición del período de un péndulo en función de su longitud; etc.

- ✓ interesa estudiar si dos variables mantienen una correlación (causal o no) y cuál es el grado de esta vinculación o dependencia. Por ejemplo: estudio de la relación entre el peso y la altura de personas; relación de consumo de gas natural y la temperatura; relación entre la velocidad máxima que alcanza un velero y su extensión desde proa a popa; etc.

Se trata de que la información que se represente quede expuesta de una manera lo suficientemente clara y explícita como para que la representación gráfica "hable por sí sola". Lo importante es que un gráfico debe servir para un posterior tratamiento de los datos, que lleve a inferir las leyes subyacentes en ellos y ahondar así en las posibles implicaciones y generalizaciones de los resultados obtenidos en los estudios o experimentos.

Un gráfico debe construirse sobre la base de una elección adecuada tanto de las variables como de las escalas de los ejes. Comentaremos diversas opciones que se presentan y sobre algunos métodos numéricos de utilidad para el tratamiento general de los datos.

Elección de las variables

Es común al estudiar un sistema, tratar de investigar la variación o dependencia de un dado atributo del mismo (genéricamente digamos la variable *Y*) como función de otra variable del sistema, que llamaremos *X*, y que sospechamos esta relacionada con la variable *Y*. La Fig. 1 representa esquemáticamente un sistema bajo estudio.

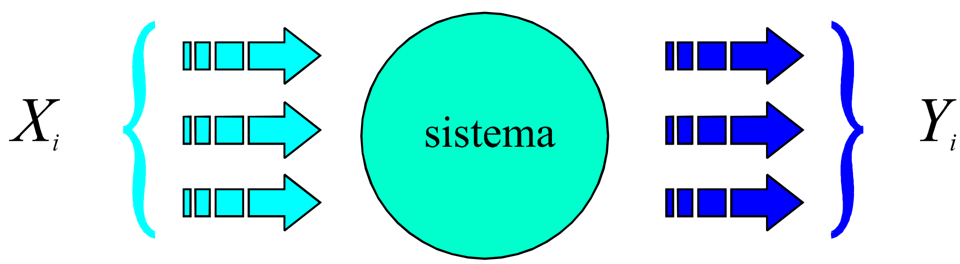


Figura 1 Representación de un sistema al que se estudia las respuestas Y_i cuando se varía el conjunto de variables X_i.

Hemos llamado X_i a las "*variables que consideramos independientes*", o sea aquellas que podemos controlar y variar. Ante los cambios de X_i, el sistema revela sus características o comportamientos a través de los cambios que sufren las variables Y_i, que pueden llamarse las "*variables de salida o dependientes*".[1,2, 3] Si deseamos estudiar un sistema, es conveniente, siempre que sea posible, aislar o controlar lo más posible las variables en estudio. Para ello es conveniente diseñar el estudio o experimento de modo tal que en lo posible solo un parámetro relevante varíe por vez, manteniendo los restantes parámetros constantes. De este modo podremos concentrarnos en la respuesta de una de las variables de salida ante las variaciones de solamente una de las variables de entrada. Siempre que esto sea posible, es muy conveniente buscar esta condición para simplificar el análisis. Afortunadamente esta es una situación es posible en algunos casos. En sistemas de mayor complejidad, en los que no podemos aislar los parámetros de modo que varíen de a uno por vez, el tipo de análisis que mostraremos puede generalizarse para tratar esos casos.[2] En lo que sigue nos apoyaremos en algunas relaciones funcionales simples con las que nos encontramos a menudo en diversos estudios y las usaremos para ejemplificar las ideas básicas.

Relación lineal

Una relación lineal entre las variables X e Y

$$Y = a \cdot X + b \tag{1}$$

es la más simple de todas. La representación gráfica de $Y(X)$ arrojaría una línea recta, de pendiente a y que corta al eje vertical en b (ordenada del origen o intersección con el eje y). Es importante notar que una recta es la forma geométrica más simple en dos dimensiones. Al mismo tiempo, una relación lineal entre dos variables es muy fácil de identificada a simple vista, y no sería una exageración afirmar que éste es el único caso en que esta discriminación puede hacerse a simple vista. Entre una recta y una curva nuestro ojo puede con frecuencia notar la diferencia, pero difícilmente discriminará a la función que define la curva, es decir a simple vista es muy difícil de saber si las variables presentan un relación potencial, exponencial o de otro tipo.[4]

En la Fig. 2 están representadas dos series de datos. Para inferir cualitativamente cuál de las series puede aproximarse mejor por una relación lineal entre las variables X e Y, es útil la siguiente regla práctica: llevemos el papel hasta el nivel de nuestros ojos (podemos cerrar uno como cuando hacemos puntería) y veamos si los puntos se ven alineados. Este tipo de toma de decisión no debe desdeñarse en el momento de analizar datos experimentales. La decisión de aceptar o no una relación lineal entre las variables debe ser tomada por el investigador, ya sea se espere o no una vinculación lineal entre las variables en juego. Una vez que decidimos que los datos “caen sobre una recta”, recién podremos estimar los parámetros (pendiente y ordenada al origen) de la *mejor recta* que aproxime la relación funcional: o bien podemos dibujar esa mejor recta y definirle los valores de la pendiente y la ordenada al origen, o usar métodos numéricos más generales para encontrarlos, como veremos más adelante.

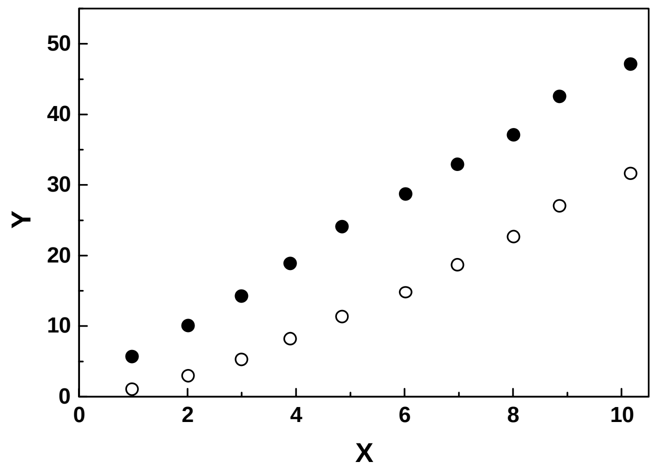


Figura 2 Representación de dos series de datos. ¿Cuál aproxima mejor una relación lineal entre las variables *X* e *Y*?

Función potencial

Supongamos que medimos pares de valores (X,Y) y tenemos conocimiento que la relación funcional que los vincula es del tipo

$$Y = aX^b \qquad (2)$$

donde a y b son constantes. Esta forma funcional potencial es muy común en las ciencias puesto que sirve como aproximación del comportamiento en una gran variedad de casos.

La constante b suele llamarse “exponente de escala” y define la escala de variación de Y según varía X. Esto es, si X se multiplica por un factor f, Y cambiará consecuentemente f^b veces.

El significado de la constante a es el de representar el valor que toma Y cuando X vale la unidad. La dimensión de a es tal que da homogeneidad dimensional a la ecuación.

Lectura de ecuaciones: Algunas investigaciones muestran que la masa de los dinosaurios M estaba bien correlacionada con la longitud l de los animales medida desde la cabeza a la cola, según[4]

$$M = M_0 l^3$$

Leamos esta ecuación. El significado de M_0 es que representa la masa de un dinosaurio de "largo unidad". Por tanto, si la unidad elegida para la longitud es el metro y para la masa el kg, M_0 representa cuántos kilogramos pesaba un animal de largo igual a 1 m. La *unidad* de M_0 será tal que se igualen las unidades de los dos miembros de la ecuación. En este caso, M_0 tendrá la unidad kg/m^3. Sin embargo, M_0 no es la densidad de los animales, a pesar de su unidad, puesto que l^3 no es el volumen. Notemos que el valor de M_0 cambiará si se eligen otras unidades de medición. Por ejemplo, si el peso se midiera en gramos (g) y la longitud en centímetros (cm), M_0 adoptaría un nuevo valor, que sería M_0^* (g / cm^3) $=10^{-3} M_0(kg/m^3)$, a lo que se llega tras pasar los kilogramos a gramos y los metros a centímetros.

De manera más general, y sin recurrir a unidades particulares, podemos analizar cuál es la *dimensión* de M_0. Si usamos corchetes [...] para representar la dimensión de una cantidad, entonces $[M_0] = \frac{[M]}{[l]^3}$. Escribamos esta relación dimensional en términos de las dimensiones fundamentales masa, longitud y tiempo, a las que llamaremos M, L y T, respectivamente. Dado que $[M] = [m] = M$, resulta, luego de simplificar: $[M_0] = M / L^3$.

Este tipo de análisis puede usarse como prueba de consistencia de una fórmula complicada, o bien para determinar la dimensión de alguna variable introducida en un problema particular.

Si representamos datos medidos de Y en función de X, relacionados por una expresión como (2), lo que obtenemos en el caso $b \neq 1$ es una curva. De nuestro análisis cualitativo del gráfico observaremos una curva "cóncava hacia arriba" si $b > 1$, mientras que si $b < 1$, la curva se verá "cóncava hacia abajo". Lo que cualquiera de los casos precedentes significa es que una variación de la variable X a un dado ritmo, hace que la variable Y cambie a un ritmo distinto: más rápido si $b > 1$, más lento si $b < 1$. Esta observación cualitativa (en términos de "más rápido" o "más lento", bien puede ser buena en una gran variedad de casos de interés en el laboratorio cuando estemos interesados en la descripción general (tendencia) de algún fenómeno.

Transformación de variables

Si en la Ec. (2) transformamos las variables haciendo el cambio (suponiendo que conocemos el exponente b):

$$X^* = X^b \qquad\qquad Y^* = Y$$

y representamos las nuevas variables $(X^*, Y^*) = (X^b, Y)$, lo que logramos es una relación lineal entre las *variables transformadas* o *pseudovariables* y decimos que hemos *linealizado* la representación gráfica. En este caso hemos transformamos la variable X, pero bien podríamos haber optado por el cambio en la variable dependiente, o sea,

$$X^* = X \qquad\qquad Y^* = Y^{1/b}$$

y también habríamos obtenido una relación lineal entre las nuevas variables representadas $(X^*, Y^*) = (X, Y^{1/b})$.

Está claro que lo anterior es inmediato de realizar si conocemos el valor del exponente b. Además, observamos que un gráfico linealizado nos da el valor de la constante a [ver Ec. (2)] si evaluamos la pendiente de la recta que resulta.

📖 Se mide el período T de un péndulo simple para distintas longitudes L. En el caso de pequeñas amplitudes de oscilación, ambas variables están relacionadas por

$$T = 2\pi\sqrt{\frac{L}{g}}$$

donde *g* es la aceleración de la gravedad. La relación es del tipo potencial:

$$T = aL^b$$

con $a = \frac{2\pi}{\sqrt{g}}$ y $b = \frac{1}{2}$.

De un gráfico de T en función de L^b evaluamos la constante a, con lo que podemos obtener el valor de la aceleración gravitatoria *g* indirectamente. Es de esperar que el gráfico resulte en una recta que pase por el origen de coordenadas, dado que un péndulo de longitud nula tiene que tener un período de oscilación nulo.

En el caso más general, supongamos que no conocemos a a ni a b, y que ambas constantes deben encontrarse como resultado de la investigación. Entonces, ¿cómo procedemos?

Para facilitar la tarea de encontrar tanto el exponente de escala b como la constante a, es conveniente representar $\log(Y)$ en función de $\log(X)$. Esto queda claro si transformamos nuestra ecuación original más general $Y = aX^b$, sacándole el logaritmo a ambos miembros

$$\log(Y) = \log(aX^b) \tag{3}$$

$$\log(Y) = \log(a) + \log(X^b) \tag{4}$$

$$\log(Y) = \log(a) + b\log(X) \tag{5}$$

Comparando esta última expresión con un gráfico de $\log(Y)$ en función de $\log(X)$ podremos ver que la ecuación representa una recta que tiene pendiente b y ordenada al origen $\log(a)$.

Este tipo de representación gráfica es extremadamente útil cuando se analizan ecuaciones algebraicas, se estudian correlaciones, leyes de crecimiento, etc.

Elección de las escalas

Hemos visto cómo elegir las nuevas variables con el fin de llevar la representación gráfica a una representación lineal. Lo que hemos propuesto es la transformación de las variables y la representación de las nuevas. Una manera alternativa de análisis es recurrir a gráficos en los que sus ejes tengan escalas logarítmicas.

Retomando el ejemplo del caso de variables *X*, *Y* relacionadas por la función potencial $Y = aX^b$, en vez de recurrir a un gráfico de variables transformadas [log (*X*), log(*Y*)], podemos representar directamente los pares de valores (*X*, *Y*) en un gráfico donde sus dos ejes tengan escalas logarítmicas. La Figuras 3 y 4 se ejemplifica este método.

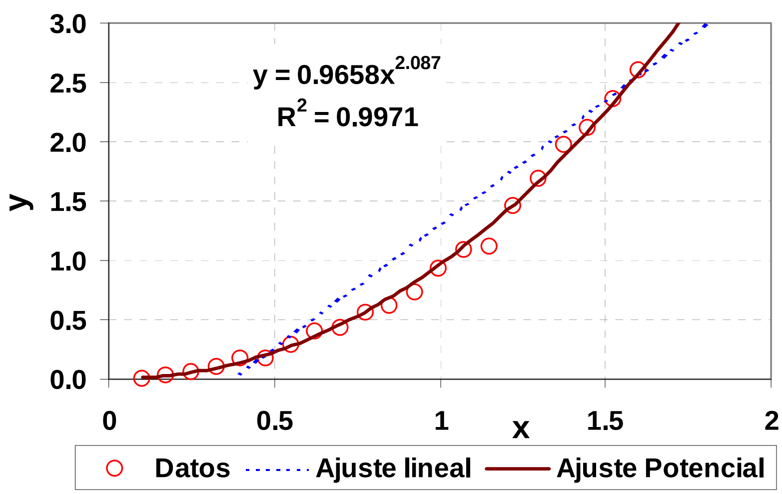


Figura 3 Representación de dos series de datos con dependencia potencial en escala lineal. La lineal de puntos es un ajuste lineal a los datos, mientras que la curva continua llena es un ajuste potencial (Y=$a.x^b$).

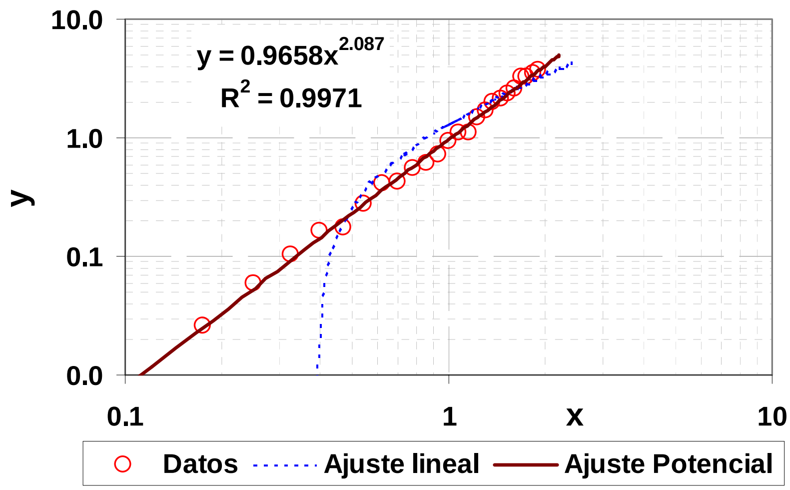


Figura 4 Ejemplo de un gráfico con escalas logarítmicas. Los datos representados en este gráfico son exactamente los mismos que los de la Figura 3, solo las escalas son distintas. En este caso ambas escalas son logarítmicas. Nótese que la recta, representada por la línea de puntos, en esta escala no es más una recta, mientras que la curva potencias, curva en línea llena, en esta representación se ve como una recta.

Un *gráfico doble-logarítmico* como el de la Fig. 4 también es llamado *gráfico log–log*. Nótese que en escala logarítmica, las décadas son equidistantes, o sea la distancia entre 0.1 y 1 es igual a la de 1 hasta 10 y así sucesivamente. Esto es muy diferente a lo que ocurre en las escalas lineales o normales, donde la distancia entre 0 a 1 es igual a la de 1 a 2 y así sucesivamente, como el la figura 3.

Observando la Fig. 4 podemos darnos cuenta que las escalas logarítmicas son “más democráticas” que las lineales,[3] puesto que dejan ocupar el mismo espacio en el gráfico a los intervalos entre décadas entre valores “pequeños” que el espacio ocupado por los intervalos entre décadas entre valores “grandes”; podemos ver, por ejemplo, que el lugar reservado para los valores entre 10^{-1} y 1 es idéntico al reservado para el intervalo 1 y 10.

Si las variables X e Y se representan ambas en escalas logarítmicas, la función potencial de la Ec. (2) quedará representada por una recta cuya pendiente es b y cuya ordenada al origen $ord = \log(a)$, por lo que $a = 10^{ord}$, como lo ilustra la figura 4.

A su vez, si los datos (X, Y) que suponemos tiene una dependencia potencial, son representados en un gráfico doble-logarítmico, los mismo mostrarán una tendencia lineal. Esto nos permite inferir si un conjunto de datos, al ser graficados en un escala doble log, o log-log, muestran una dependencia lineal, esto sugiere que la relación funcional entre los mismos es potencial, o sea que $Y \propto X^b$. Por lo tanto el grafico nos permite descubrir en este caso la ley subyacente, o sea la relación funcional entre las variables X e Y. Para calcular directamente del gráfico el valor del exponente b, hay que contar cuántas décadas varía Y cuando X varía una. En el ejemplo de la Fig. 4, la línea llena tiene pendiente $b \approx 2$, ya que por una década de variación de X , tenemos dos décadas de variación de Y.

Esta representación puede hacerse sobre papeles especialmente diseñados (papel logarítmico) que se consigue en las librerías. Con las ventajas que ofrecen hoy en día los programas de computadora, este tipo de representación puede realizarse de manera inmediata para sacar mayor provecho al análisis de los datos experimentales. Muchos programas de análisis de datos o planillas de cálculo, tales como Excel, QuatroPro, Origin, etc., permiten realizar estos cambios muy fácilmente. Una vez realizado el gráfico en escala lineal, picando o activando con el *mouse* los ejes coordenados, se abre un sub-menú que permite variar la escala de los ejes (lineal, logarítmica, etc.).

La ley exponencial

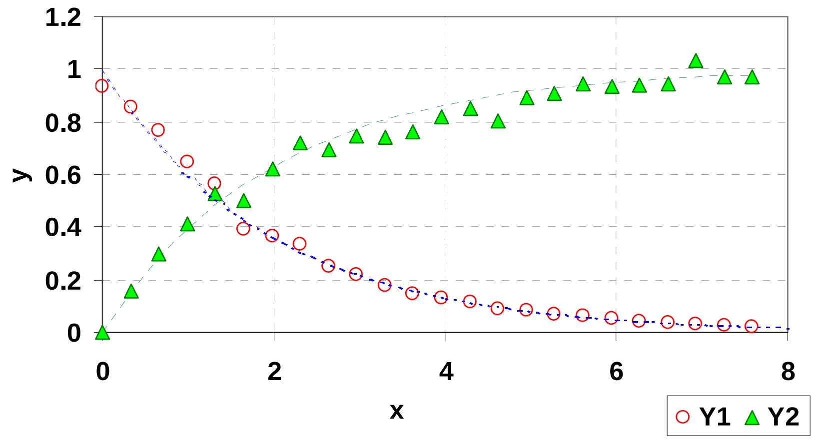
Un caso particular de mucho interés por su aplicación en muchos problemas físicos, biológicos, de ingeniería, financieros y económicos, es el de una relación exponencial entre dos variables. Para fijar ideas supongamos que estamos considerando dos variables, Y_1 e Y_2, como función del tiempo t. Si las relaciones entre estas variables son:

$$Y_1(t) = Ae^{-\lambda_1 t} \tag{6}$$

y

$$Y_2(t) = A(1 - e^{-\lambda_2 t}) \tag{7}$$

sus representaciones gráficas lucirán en escala semilogarítmica como se muestra en la Fig. 5.



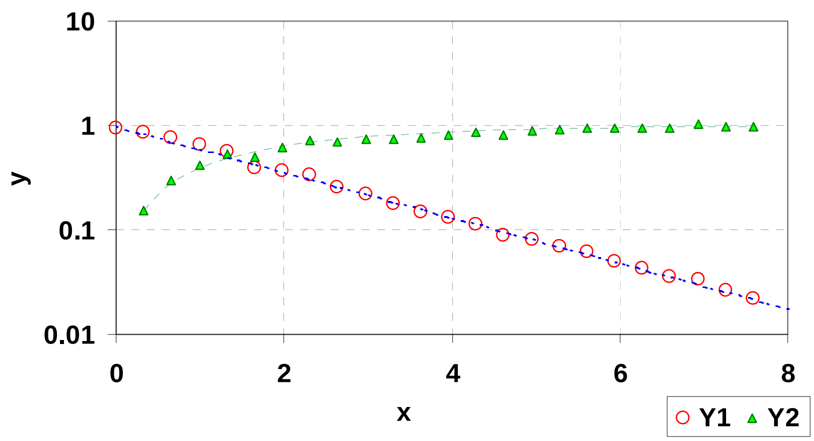


Figura 5 Representación en escala de las funciones (6) y (7) en escala normal o lineal (panel superior) y en escala semilogarítmica (panel inferior). Nótese que solo la expresión (6) se linealiza en escala semilogarítmica.

Es fácil notar que, si bien la primera de estas relaciones (Y_1) se "linealiza" en escala semilogarítmica, la segunda (Y_2) no lo hace. En este caso, es conveniente recordar que la derivada de ambas expresiones sí tienen una relación funcional simple, a saber:

$$\frac{dY_1(t)}{dt} = -A\lambda_1 e^{-\lambda_1 t} = -\lambda_1 Y_1(t) \tag{8}$$

y

$$\frac{dY_2(t)}{dt} = A\lambda_2 e^{-\lambda_2 t} = \lambda_2 (A - Y_2(t)) \tag{9}$$

Por lo tanto, usando una representación de la *derivada en función de la variable dependiente* (dY_1/dt o dY_2/dt) en función de (Y_1 o Y_2) es cuando obtenemos una recta. De los valores de la pendiente y la ordenada al origen de estas rectas (8) y (9), tenemos información sobre este tipo de relación, puesto que de ellos extraemos los parámetros λ_1 y λ_2. En la Fig. 6 se muestran las mismas funciones que en la Fig. 5 usando la representación propuesta. Es claro que esta alternativa es muy útil para estos problemas.

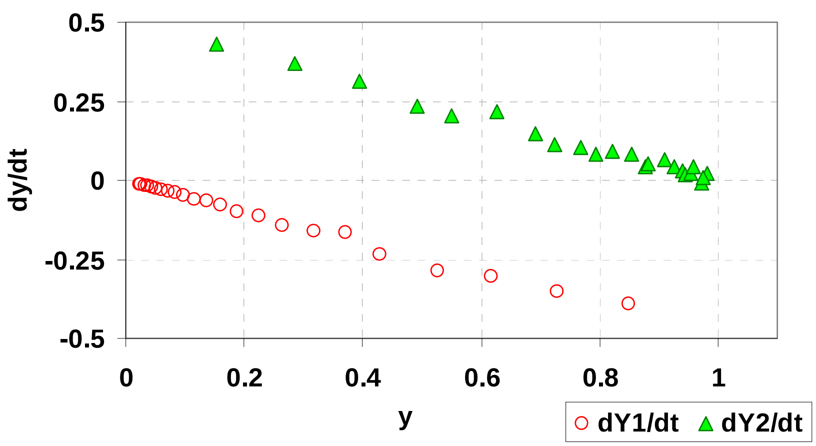


Figura 6 Representación en escala lineal de las derivadas *dY1/dt* y *dY2/dt* en función de las variables dependientes *Y1* e Y2 respectivamente. En este ejemplo, $\lambda_1 = \lambda_2$, por lo tanto los datos se alienan en dos rectas paralelas.

Una dificultad de esta representación es que requiere conocer la derivada de la función en cuestión y que para hacerlo debemos usar un procedimiento numérico. Si disponemos de mediciones de Y_1 e Y_2 en función de t lo que hacemos es aproximar la derivada calculando las diferencias finitas usando pares de datos consecutivos:

$$\frac{dY(t)}{dt} \approx \frac{Y_{i+1} - Y_i}{t_{i+1} - t_i} \tag{9}$$

Sin embargo, como los datos tienen errores, la diferencia ($Y_{i+1} - Y_i$) puede ser en algunos casos menor que el error de medición, y en tal caso el valor obtenido con (9) presenta mucha dispersión. Una manera de mejorar la estimación de la derivada de datos experimentales consiste en usar un grupo de datos que estén en un intervalo donde *a priori* no se espere mucha variación en la derivada. Usando este grupo de valores elegidos (Y_i, t_i) aproximamos una recta que pase por ellos, cuya pendiente m tomamos como una estimación de la pendiente de la curva en el entorno de esos datos, o sea, hacemos una estimación local de la derivada dY/dt usando un grupo de valores en vez de usar pares consecutivos. La función pendiente (*Slope*) es una función muy común en muchas planillas de calculo. El gráfico que hacemos finalmente es uno de m en función de Y. La mayoría de la hojas de cálculo usan este procedimiento para el cálculo de la derivada de una función representada por un conjunto de datos.

Diseño de gráficos

Los programas de representación gráfica disponibles en las computadoras incluyen entre sus opciones el diseño de gráficos usando los distintos tipos de escalas descriptas en este capítulo. Algunas sugerencia útiles de seguir son las siguientes:

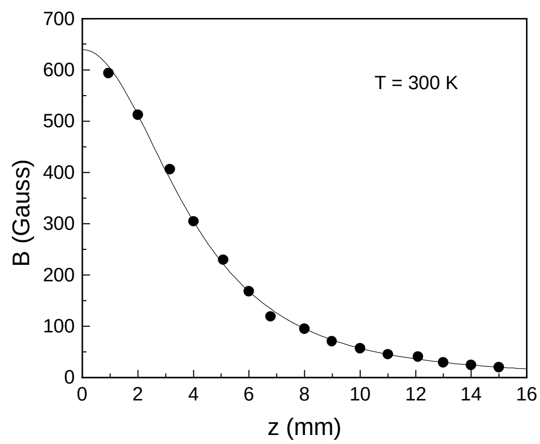
- ✓ identificación de los ejes con *rótulos* bien ubicados que digan qué variables se representan y en qué *unidades* se miden,

- ✓ uso de *símbolos* que ubiquen los datos (cuadrados, círculos, rombos, etc.), en lo posible con sus *incertidumbres* (en la forma de barras que indiquen el intervalo de incertidumbre); que haya una diferenciación de distintas series de datos cuando se presenten varios resultados, para lo que es recomendable el uso de diferentes símbolos,

- ✓ inclusión de un epígrafe, que es un texto descriptivo de lo que está representado en el gráfico y que además puede aportar alguna información adicional,

- ✓ carteles interiores al gráfico, con información complementaria relevante para entender en qué contexto se muestran los datos o sobre las condiciones experimentales particulares bajo las que se los han obtenido,

✓ una clara diferenciación entre los símbolos que se usan para indicar los datos a analizar, por ejemplo resultados experimento, y los que corresponde a una teoría o modelo propuesto que se propone para explicar los datos (por ejemplo, usando líneas continuas). Por regla general se usan símbolos (cuadrados, círculos, etc.) para representar datos reales o los resultados de una medición o experimento, mientras que se usan líneas continuas para representar el modelo o la teoría explicativa que se usa para interpretar los mismos. En la figura 7 se ilustra este proceder.



Campo magnético axial de un imán medido con una sonda de efecto Hall. La línea es un ajuste de los datos.

Figura 7 Ejemplo de gráfico y epígrafe o leyenda. Los círculos llenos corresponden a los resultados medidos y la línea continua es un modelo que pretende explicar dichos datos.

La navaja de Occam – Criterio de Parsimonia

"*Las descripciones deben mantenerse lo más simples posibles hasta el momento en que se demuestre que resultan inadecuadas*"

La navaja de Occam establece que al elaborar una teoría o explicación de un fenómeno, uno no debe hacer más suposiciones que las mínimas necesarias. Este principio filosófico se conoce también como *criterio de parsimonia*. Estas ideas están subyacentes en todo el pensamiento científico y filosófico. Es además muy útil a la hora de elaborar modelos explicativos.

Si se puede explicar el comportamiento de un fenómeno con pocas variables explicativas y si la teoría explicativa pertinente no es lo suficientemente fuerte para sugerir otras variables que deban ser incluidas, ¿porqué introducir más variables? Por ejemplo, si un fenómeno se puede explicar por una relación lineal, ¿por qué usar un polinomio de 5 grado?. Si los datos se ajustan por una recta del tipo $y=ax+b$, y b es cercano a cero, siempre es conveniente preguntarse sui nuestras datos pueden efectivamente explicarse por una relación de tipo $y=a.x$, nótese que esta última expresión solo tiene 1 parámetro libre (a) mientras que la anterior tenia dos (a y b), por lo tanto la última es 50% más simple y económica. Lógicamente, si a ajustar los datos con $y=a.x$ obtenemos un mal ajuste y con $y=ax+b$, el ajuste ex bueno, en ese caso optamos por la expresión con los dos parámetros.

"Lo bueno si breve dos veces bueno"

Ejemplo: Al estudiar la relación entre el estiramiento de un resorte en función de la fuerza aplicada, ver figura 8, se encuentra que la recta que mejor ajusta los datos es:

$$F(N) = 90.25 \cdot \Delta x(m) - 0.04 \qquad \text{con} \qquad R^2 = 0.991$$

ΔX(m)	F(N)
0	0
0.02	1.65
0.033	3.2
0.045	3.6
0.06	5.5
0.075	7
0.091	8

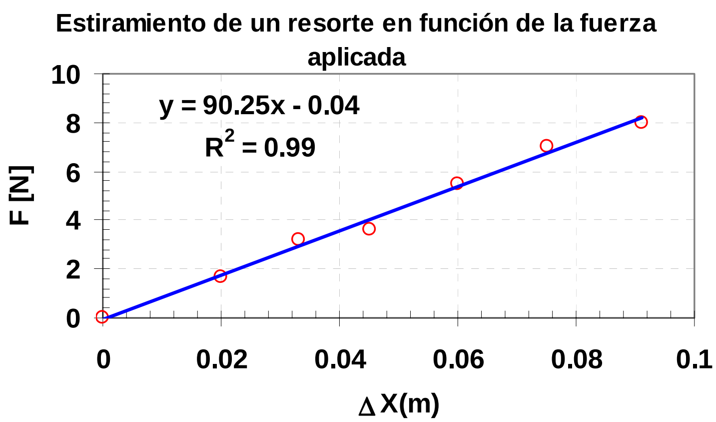


Figura 8 Ejemplo de datos y gráfico resultado de estudiar la relación entre el estiramiento de un resorte en función de la fuerza aplicada

Referencias

1. S. Gil y E. Rodríguez, *Física re-Creativa*, Prentice Hall, Buenos Aires, 2001.; http://www.fisicarecreativa.com
2. D. C. Baird, *Experimentación*, 2ª ed., Prentice-Hall Hispanoamericana S.A., México, 1991.
3. Christopher Deacon, "The importance of graphs in undergraduate physics," *Phys. Teach.* **37**, 270, 1999.
4. E. Martínez, *Logarithmic Park*, Instituto Balseiro, Bariloche, 1997; http://cabbat1.cnea.gov.ar/apfa/apfa.htm

Ejercicios

I) Proponga el tipo de gráfico (lineal, log–log, semi–log) y las correspondientes varibles o pseudovariables que permitan linealizar la representación gráfica de cada una de las siguientes funciones. Indique en cada caso el procedimiento a seguir para encontrar, a partir del gráfico, los valores de las constantes *a* y *b*.

1) $y = ax^2 + b$
2) $y = ax^b$
3) $y = \frac{a}{x} + b$
4) $\frac{1}{y} = \frac{1}{x} + a$
5) $y^2 = ax^3 + b$
6) $y = a\exp(-bx)$
7) $y = a\exp(-\frac{b}{x})$
8) $y = a\log(bx)$

II) Considere la expresión $\frac{1}{x} + \frac{1}{y} = \frac{1}{a}$. Demuestre que hay dos maneras de linealizar la representación gráfica:

i) representando $\frac{1}{y}$ en función de $\frac{1}{x}$,

ii) representando el producto xy en función de la suma $x + y$.

- Indique cómo se puede obtener el valor de la constante *a* de cada gráfico